Self-Publishing and Bibliographic Control

Robert P. Holley, Wayne State University

INTRODUCTION
This chapter will investigate the importance of bibliographic control for self-publishing, the current status of bibliographic control, and possible future steps. The number of self-published books has rapidly increased in the United States, accounting for at least 50% of all titles published, though accurate figure are difficult to determine in part because of the lack of bibliographic control in contrast with commercially published books (Bowker). Of course, though their numbers are greater, the sales volume of self-published materials is much less than that of commercial publications. Issues of their quality partially explain this disparity; but difficulties in discovery, distribution, inventory control, and library/bookstore processing limit the possibility for potentially popular self-published books to meet their market potential. More effective bibliographic control including both traditional cataloging and accurate metadata would help overcome these constraints and level the playing field with commercially published items.

THE IMPORTANCE OF BIBLIOGRAPHIC CONTROL
Bibliographic control has great importance for the success of publications for two principal reasons: discovery and distribution/handling. Librarians, publishers, and booksellers have come to depend upon sophisticated mechanisms that have developed over decades, if not centuries, to publicize, distribute, and find trade and university press publications. These mechanisms have become second nature so that many may not realize
how their minimal support for self-published books hampers their distribution to libraries and bookstores and thus lessens their availability to readers.

Discovery is an extremely important part of the publishing cycle. Librarians, bookstore owners, and readers cannot acquire materials that they don’t know exist. Cataloging and other metadata play a key role in helping all these groups find publications. (In the following discussion, the term cataloging also includes other metadata since cataloging has been the traditional way libraries deal with bibliographic control.) Finding a known item is the least difficult task as long as the searcher has enough bibliographic information about a title, which can now be only a few words. The availability of keyword searching has eliminated the need for the perfect citation required in the print age. For known-item self-published materials, querying Amazon, WorldCat, Goodreads, Smashwords, AddAll, and similar bibliographic databases or even searching with Google (arguably the most comprehensive search engine) should be able to find the item. If these strategies don’t work, the item might as well not exist since no practical way exists to get a copy.

Cataloging becomes more important when the searcher wants to expand a search to discover items of potential interest. Traditional library records including both bibliographic and authority files can do, at least in principle, all the following things:

- Find all items where an author has some sort of responsibility
- Identify variant forms of the author’s name and any pseudonyms
- Discover coauthors whose works may also be of interest
- Identify title changes, variant forms of the title, and the title in foreign languages
- Determine other items in a series or set whether by the same author or different authors
- Find publications connected with a conference even when the conference changes name or location
- Find all items published by a specific publisher, in a specific location, or even during a specific date range

All these characteristics are reasonably objective and do not depend upon the cataloger’s judgment in the way that those in the next section do.
The second set of discovery characteristics deals with the subject content and genre of the work. Keyword searching, while extremely powerful, has its limitations. It does not handle well character strings with radically different meanings. Searching for “mars” can retrieve materials about a Roman god, a planet, a candy bar, and the month of March in French. It also has difficulty with synonyms, regional differences in meaning, and foreign languages. Traditional cataloging and other well-constructed metadata work to solve the problems with these inconsistencies. Even more importantly, traditional cataloging attempts to describe the content through systems like Library of Congress (LC) subject headings, Sears subject headings, and the various systems used in databases, sometimes with the help of their accompanying thesauri. Classification deals with the same issue by grouping similar materials together, often in some sort of hierarchical arrangement. The major classification systems, Library of Congress Classification, the Dewey Decimal Classification, and the Universal Decimal Classifications, are also language independent. In addition, cataloging often identifies the genre of fictional works. Traditional cataloging, as imperfect as it may be, provides access to traditionally published works that is often lacking for self-published materials. A discussion of ways in which potential readers can discover self-published works appears later in this chapter, but one nonbibliographic tool should be noted here that is effective for finding both trade and self-published books, namely, the feature in Amazon, “What Other Items Do Customers Buy after Viewing This Item?” This creative innovation depends upon the collective behavior of Amazon buyers and not on any efforts at bibliographic control.

The second goal of bibliographic control is to control the item as it wends its way through the various steps in the publication, distribution, and use cycle. The various elements described above all play an important role at times; but one feature not yet discussed is critical for the effective management of books, that is, the International Standard Book Number (ISBN). The ISBN was designed to serve as a unique identifier for a specific manifestation of any title and to distinguish among multiple editions and multiple formats. Gordon Foster developed the 9-digit Standard Book Number in 1965. The United States adopted the modified 10-digit ISBN in 1974. In 2007, the ISBN expanded to 13 digits while still maintaining compatibility with the earlier versions (“International Standard Book Number,” 2015).
The ISBN, normally acquired by the publisher or author, is assigned in the United States by R. R. Bower (Bowker). With its goal of providing a unique identifier for each manifestation of the work, the ISBN could, at least in principle, be used to control the item without any additional bibliographic information. Since the ISBN is often represented as a bar code, scanners can automatically read the ISBN for tracking purposes. As a unique identifier, the ISBN allows publishers, library jobbers, librarians, and out-of-print vendors to identify and track a book. Similarly, the ISBN provides the most efficient way to search bibliographic files of all types to find a known item among all the possibilities presented by the other types of information such as author, title, and series. In fact, the ISBN may be the only effective way to discover information about bibliographically complex works such as the Bible and works by Shakespeare.

For libraries, bibliographic control is especially important because they need bibliographic records that are as complete as possible in their online finding tools for both user access and internal functions. Libraries expect to acquire these bibliographic records as cheaply as possible as part of the book’s purchase price. If the library has to create a record for an item, called original cataloging, this greatly increases the amount of staff time needed even for a brief record. An accurate cost for original cataloging is difficult to determine; but PrairieCat Support, part of a consortium of 127 libraries in Illinois, charges $10 per record for basic original cataloging of standard materials (PrairieCat). A much older article by George D. Harris gave the cost of original cataloging to be as high as $32 but cautioned that cataloging costs mean something different to each library (Harris, 1989). The MARC format was created in the 1960s explicitly to share bibliographic records among libraries, bookstores, and other bibliographic agencies as a way to cut costs (Avram, 2009). OCLC and then later RLG were formed to make cataloging more economical by sharing records created by their members through their extensive online databases. Today, all but the largest libraries can easily acquire most bibliographic records for trade publications from bibliographic utilities like OCLC or from vendors who provide them electronically when the book is ordered or received. One last possibility, mostly for very small libraries, is for libraries to transcribe the cataloging found in the book since most trade publishers include this cataloging from the Library of Congress Cataloging in Publication (CIP) program.
SELF-PUBLISHED MATERIALS AND BIBLIOGRAPHIC CONTROL

The lack of bibliographic control for self-published materials is one of the key obstacles standing in the way of broader library discovery and purchase of self-published books. For American libraries, cataloging records produced by the Library of Congress are considered the best and serve as the gold standard for bibliographic control that other catalog record producers should take as their model. Library of Congress produces these records at several levels, but the full records that most small to mid-size libraries are accustomed to using offer a broad range of information including authors, titles, series, subject headings, and the classification number of the Library of Congress Classification and often the Dewey Decimal Classification. Some records include alternate call numbers and special subject headings for children. LC distributes these records in machine-readable formats relatively inexpensively. The 2015 subscription price for an estimated 238,000 new records was $19,155; the complete retrospective file of 12,668,850 records costs only $27,750 (Library of Congress, Cataloging Distribution Service, 2015). Since these records are not copyrighted for use within the United States, they can then be redistributed without charge (Library of Congress, Cataloging Distribution Service).

Publishers, book vendors, database providers, bibliographic utilities, and integrated library system companies have sophisticated software to manipulate these records. As stated above, many book vendors provide these MARC records to libraries at no or little cost when libraries purchase materials from them; or libraries can acquire them at a minimal cost from the OCLC bibliographic database and other sources.

The Library of Congress has two major streams for creating bibliographic records—the Cataloging in Publication program for mainstream American imprints and the results of its cataloging items for its own collection. Both streams mostly or completely exclude self-published materials, a fact that greatly hinders their discoverability and decreases their desirability for libraries. CIP explicitly excludes self-published materials because they are not from a recognized publisher (“Making cataloging hum,” 1996). In a similar fashion, LC does not collect many self-published materials for its collection because they remain tainted by the past history of vanity presses. The information section on LC Card Numbers specifically states: “Please note that many collection policy statements explicitly treat
the selection of self-published and vanity press materials.” For example, the LC policy for Literature and Language states: “Vanity press and self-published works are not collected, although self-published works of quality may be collected in areas where self-publishing is an important part of the publishing spectrum (e.g., poetry, African American literature).” LC does collect self-published materials in areas such as genealogy where self-publishing has more importance (Library of Congress, Cataloging and Acquisitions).

Libraries and vendors also catalog their own materials and provide the second best way for self-published materials to achieve relatively good bibliographic control. Vendors need these records to control the items and offer them for sale to their customers with the secondary benefit of having a record to share with the library. When cataloging records are not available from other sources, libraries create them or have the records created for them for their integrated library systems. These records make publications available to their patrons and facilitate internal tasks such as circulation, interlibrary loan, and collection analysis. Most often, both vendors and libraries share these records by adding them to one or more of the major bibliographic databases maintained by the bibliographic utilities such as OCLC and SkyRiver. Both libraries and vendors almost always create briefer records than those provided by LC, but they are usually of sufficient quality to meet the needs of libraries without significant editing. Nonetheless, the records created by vendors and publishers provide access to only a small percentage of self-published books. The chapters in this volume by Nardini and Cutler provide additional details on vendor support including cataloging as do the chapters by Bankhead and by DeWild and Jarema on their public library pilot projects.

The most exciting news for improved bibliographic control of self-published works is the cooperative arrangement between Smashwords and OverDrive to sell packages of e-books as well as individual e-books to libraries. According to the press release, this initiative will “bring 200,000+ indie ebooks to 20,000+ public libraries” (Smashwords, 2014). This development is important first because OverDrive is the dominant player in selling e-books to public libraries and has the goal of achieving the same status with school libraries and international libraries. To quote a column from Forbes.com:
OverDrive dominates the US public library market with its digital platform... In terms of market share, OverDrive says that they serve over 90% of the 16,400 US public libraries, with a 99% renewal rate in that segment. Although right now, they serve somewhat less than 6,000 schools; ultimately, they hope to serve an equally impressive share of the 98,000 school libraries in the U.S. Internationally, the company reports it is doing business with 27,000 schools and libraries. (Seave)

Even more importantly for this chapter, OverDrive states: “When you buy MARC records with OverDrive eBooks you never pay for original cataloging, and holdings are set in WorldCat” (OverDrive). To answer concerns about whether this same policy applies to self-published materials from Smashwords, David Burleigh, director of marketing and communication at OverDrive, confirmed in an e-mail: “Yes, Smashwords titles are treated the same as other titles when it comes to MARC records. We work with OCLC and eBibliofile which provide full MARC records for a small fee, and OverDrive also provides a minimum bib record for free” (Burleigh, personal communication, January 22, 2015). This partnership should create bibliographic records, available in OCLC, for thousands of the most popular self-published books.

Other sources provide partial bibliographic control for self-published materials and, by facilitating the discovery of these items, may eventually lead to more complete bibliographic records if libraries acquire the items. Book reviews, especially those produced by sources intended for libraries, satisfy the discovery function of bibliographic control and often provide reasonably complete information to create at least a minimum-level catalog record. The chapter by Eleanor Cook in this volume discusses book reviewing sources for self-published materials. Her conclusion is that self-published materials are receiving an increasing number of reviews but that the coverage is normally limited to the most popular titles.

The online bookselling sites are also important sources for information. The most comprehensive source for books is Amazon. From the author’s personal searching, Amazon includes at least 15,000,000 ranked titles that have sold at least one copy plus many others that have never been bought. Amazon provides basic bibliographic information including
publisher, format, language, ISBN, and other details depending upon the record. The advanced search function makes it possible to search in broad subject categories. With a known item, the “what others bought” function as described above can be useful to find related materials. It is important, however, to note here that Amazon does not segregate self-published books from the much larger category of traditional publications. Furthermore, independent sellers, responsible for an estimated 40% of all content on Amazon, often create minimal records that are much less useful for both discovery and cataloging (Loten & Janofsky, 2015).

More useful for bibliographic control are the sites dedicated to helping authors create self-published materials and then distribute them. Smashwords is the most important site because of its efforts to work with libraries including offering packages and individual e-books on OverDrive as described earlier (Smashwords, 2014). In addition, Smashwords sells directly to libraries that purchase larger collections and have the necessary technology to provide digital rights management. Beyond the traditional cataloging records from OverDrive, the records on the Smashwords site normally include the category, publication date, ISBN, and subject tags. Since the authors provide these tags, their ability to accurately describe the content in terms that users, including libraries, use to search may be limited. Brief test searches reveal that any search term must include all the words in the tag to find the item. This makes discovery more difficult because some authors use longer phrases that the average searcher would not consider. On the plus side, it is possible to search tags and other data elements with Boolean operators. Finally, it is possible to use Google to search within tags, a feature that the Smashwords site does not support. These tags would have some use for cataloging by providing suggestions for subject headings or library-supplied metadata. Lulu, the second largest self-publishing site, provides similar information including ISBN, copyright holder, edition, publisher, publication date, pages, binding, and keywords.

One last source for information about self-published materials is the Books in Print database that provides entries with basic information such as author, publisher, format, publication date, price, and, of course, the ISBN. As stated above, the ISBN is more useful for control of materials than for the discovery of unknown items. R. R. Bowker is the only agency in the United States that assigns ISBNs, a fact that allows the company to produce
Books in Print. Bowker also uses its databases to gather statistics on publishing, including the number of self-published materials reported above. Nonetheless, their number is underreported since no mandate requires a publication to have an ISBN. Unfortunately, Bowker also does not automatically include all self-published materials with an ISBN in its online Books in Print database. The author “must submit . . . meta-data to Bowker on an Advanced Book Information (ABI) form” (Berinstein, 2007) and “register as a publisher” (Bowker Link). The two-step process of first registering as a publisher and then filling out a form for the individual item most likely discourages self-published authors from including their items in this source. Many of the support services for self-publishers include assigning an ISBN as part of their fee.

THE FUTURE

The future for self-published items can only improve with the increasing prominence of self-published materials and the pressure upon libraries to acquire them. The chapters in this volume give evidence of the importance of self-published materials for public libraries but less so for academic libraries. Libraries will need to be motivated to find ways to discover, acquire, and catalog these items; but the popularity of prolific authors as described by Washington in her chapter puts pressure upon public libraries to meet patron demand. As the body of bibliographic information increases, additional libraries will find it easier to make use of existing records and build upon prior efforts. Smashwords’ efforts to sell to libraries through OverDrive offer the greatest hope for increased purchasing and subsequent bibliographic control of self-published books. Their preselected packages of the most popular self-published books at a very reasonable cost should lead to the availability of an increased number of acceptable cataloging records.

David Vinjamuri, who teaches branding and social media at NYU and for ThirdWay Brand Trainers, proposes another strategy. A column in his blog on the Forbes Web site proposes a solution that hearkens back to many earlier library cooperative ventures. “It would be easy to ignore these [self-published] books if they resembled the vanity press works of previous years. Some do, but others do not. Publisher’s Weekly [sic] estimates that fifteen of the 100 bestselling books of 2012 were self-published.” The problem is discovering the good ones. “Yet if each library discovered just one
interesting book a year—and shared that result with other libraries . . . there would be 16,000 interesting books for libraries to review. If we assume that just one in one hundred of those reviewed books are ‘great’ libraries would still have discovered 160 great new books to recommend to library patrons each year” (Vinjamuri, 2013).

Most if not all libraries will be forced to confront the issue of acquiring self-published materials as their numbers and importance grow. Amazon.com, Smashwords, Lulu Enterprises, and other services to support self-publishing are inspiring an increasing number of authors each year to take the plunge. Publishers Weekly’s statistic that 15 of the 100 best sellers in 2012 were self-published shows that public demand exists for these items. At some point, a vendor, library group, or a yet unknown entrepreneur will figure out a way to market self-published materials to libraries and to provide the bibliographic control that libraries, librarians, and their patrons demand.

REFERENCES


PrairieCat. (n.d.). PrairieCat support. Retrieved from https://support.prairiecat.info/services/cataloging